

Compétences AMU dans les Grandes Masses de Données

Animation « Big Data » des Poles de Recherche
interdisciplinaires et intersectoriels de l'AMU

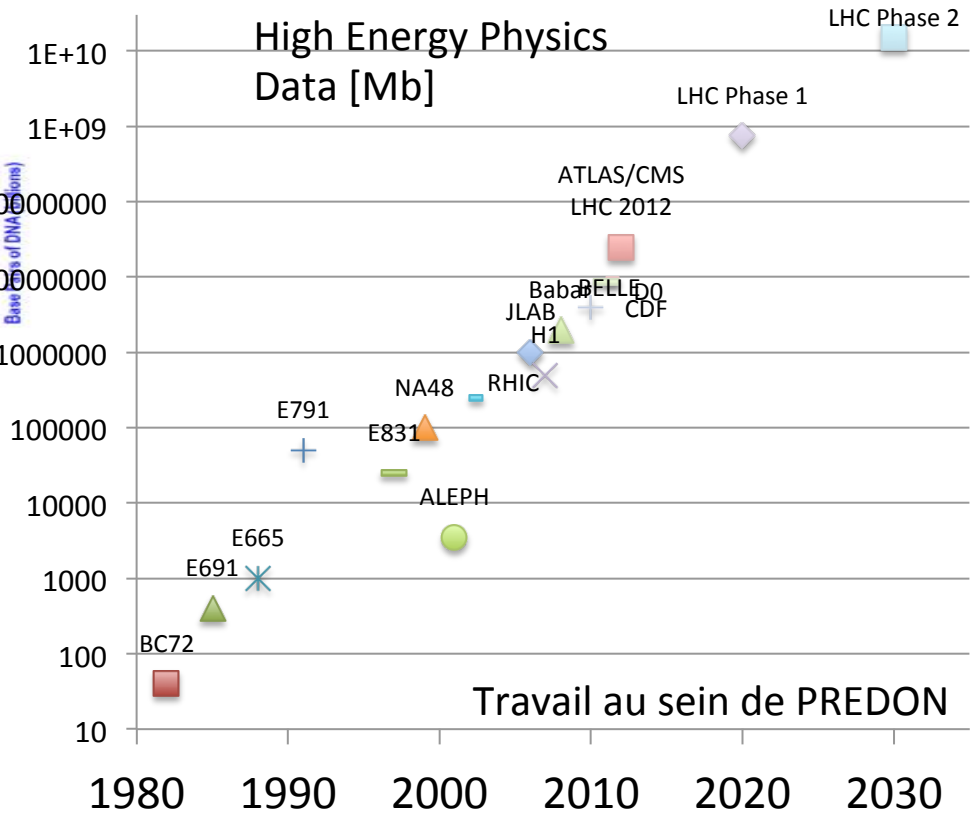
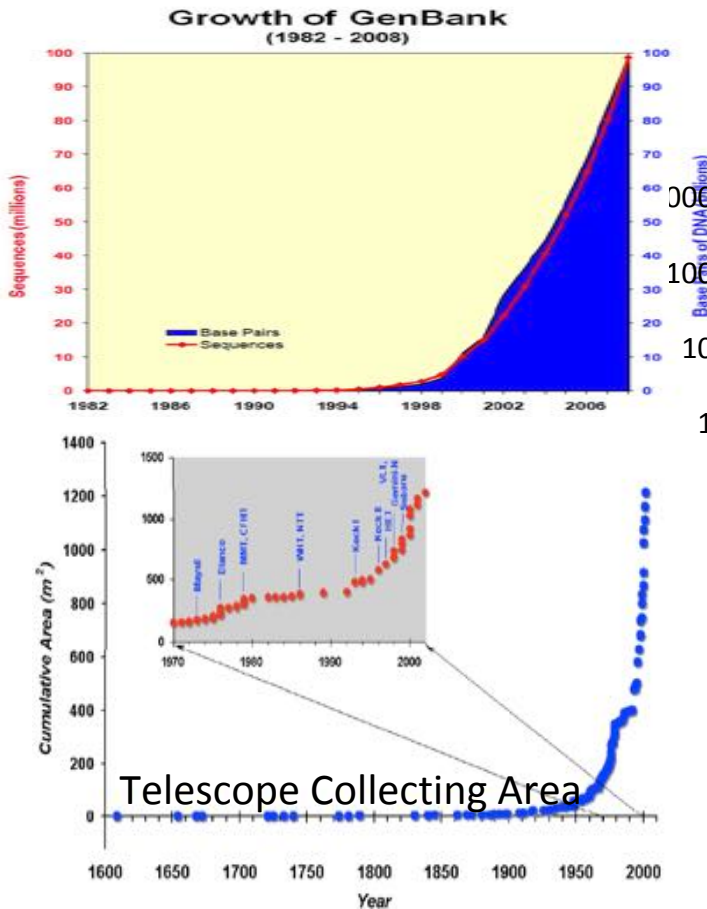
Journée: "Square Kilometer Array": des défis
Énergétique et BigData hors du commun

LAM 16 Mai 2018

(C. Diaconu)

« Big Scientific Data »

- La recherche est « digitale »
 - Augmentation dramatique de la quantité/complexité des données



Travail au sein de PREDON

Communauté « Big Data » AMU

Démarrage en 2014 de l'Animation « Big Data » AMU sous la coupe des PR2I
~30 contact personnes dans les laboratoires AMU

DRAFT V003

22/04/15 11:21 AM



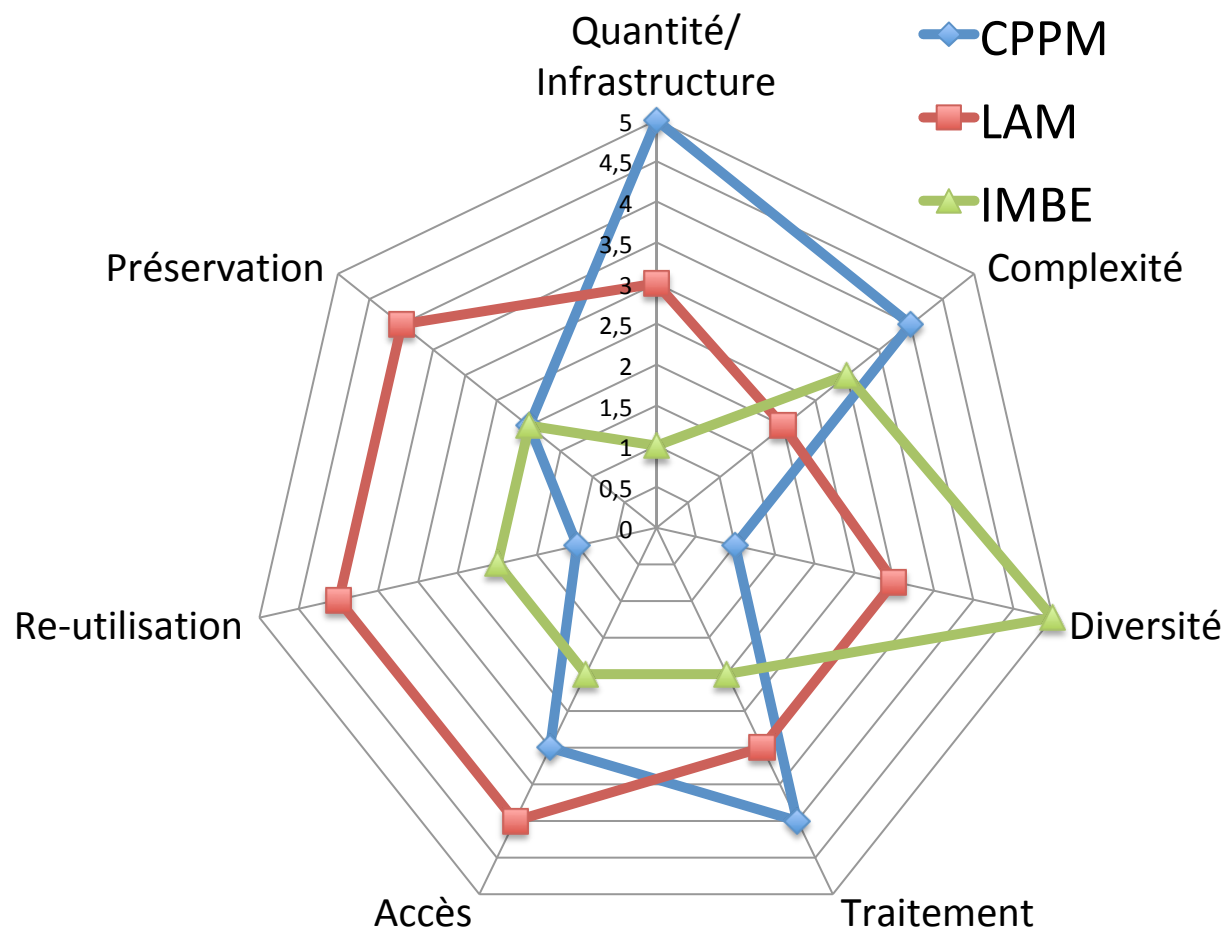
Synthèse des activités dans le domaine des grandes masses de données dans les laboratoires AMU

Introduction	3
Compétences, besoins et projets dans les laboratoires AMU	3
Méthodologies de traitement et fouille (« data mining »)	3
Mise à disposition et préservation	4
Sécurité et éthique	4
Infrastructure et technologies	5
Actions 2015/2016	5
ANNEXES	7
Annexe A: Comité de pilotage de l'animation "Big Data" au sein des PR2I AMU	8
Annexe B: Contacts dans les laboratoires	9
Annexe C: Compétences dans les laboratoires AMU (en bref)	10
Annexe D: Projets dans les laboratoires AMU (en bref)	14
Annexe E: Fiches complètes des laboratoires	18
TAGC	18
CEREGE	20
CPT	21
CRET-LOG	22
ESPACE	24
Institut Fresnel	26
GREQAM	27
I2M	29
IBDM	30
INT	31
IRSIC	32
ISm2	34
LAM	35
LBA	36
LCB	38
LIF	40
IMBE	42
LPL	44
LSIS	45
MESOCENTRE	48
SESSTIM	49
INSERM UMR S910	51
CRCM	52
Eccorev	54



Compétences et complémentarité AMU

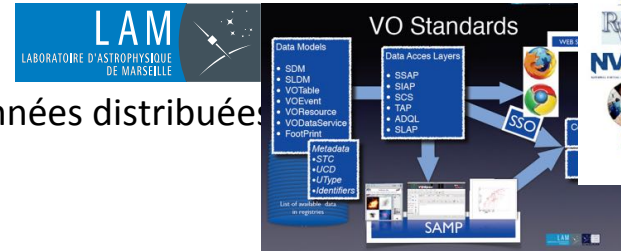
Quantité
Complexité
Diversité
Traitement
Accès
Re-utilisation
Préservation



Compétences

Astrophysique

- LAM / CeSAM : Gestion de masses de données et de données distribuées (Observatoire Virtuel)



Calcul pour le LHC

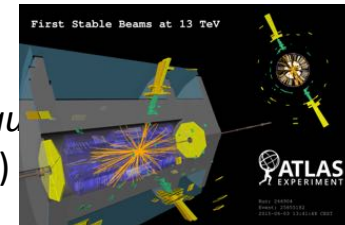
CERN EP Department - R&D on experimental technologies – Workshop, G.Stewart , A. Salzburger (CERN) <https://indico.cern.ch/event/696066/>



Apprentissage automatique

Physique des particules au LHC

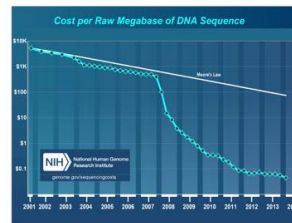
Apprentissage profond pour la recherche de phénomènes physiques nouveaux à partir des données massives collectées par l'expérience ATLAS au Grand Collisionneur de Hadrons (LHC) au CERN. (sujet de thèse – CPPM-LIS)



Génétique

Recherche de maladies rares

Professeur Christophe BEROUD
Laboratoire de Génétique Moléculaire
Hôpital TIMONE Enfants



Médecine

Methodologie pour identifier des structures dans les bigdata Reynier et al.



Infrastructures mutualisées à l'AMU

Project M3AMU funded by CPER/FEDER/CD13

- HPC (mésocentre) and HTC (grille) in one center, storage close to the computing nodes
- DIRAC interware: cloud computing, extension to GPU

CCIAM: Centre de Calcul Intensif d'Aix Marseille UMS (en cours d'installation)

- Mise en œuvre du projet M3AMU
- Participation à la politique d'établissement en matière de calcul intensif
- Mutualisation des ressources

Data Center Régional AMU: labellisation en cours

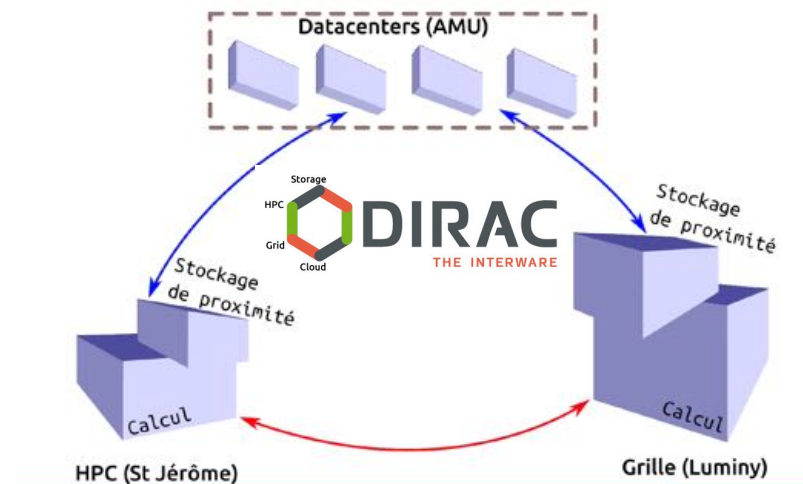
M³AMU

Mésocentre Multi-Modalités in AMU

HPC, grid, cloud, storage for scientific computing and Big Data



Choosing an infrastructure for massive data: storage & processing?



Un environnement local

- MEETUP
 - <https://www.meetup.com/fr-FR/Machine-learning-Aix-Marseille/>
 - <https://www.meetup.com/fr-FR/MongoDB-Aix-Marseille/>

et une participation nationale

- Recherche
 - Projets MASTODONS
 - Research in Big Data
 - <http://www.cnrs.fr/mi/spip.php?article985&lang=fr>
 - PEPS Astro-Informatique (http://www.cnrs.fr/mi/IMG/pdf/astro-info2018_tableau_web.pdf)
 - BIGSKYEARTH : European COST Program (European Cooperation in Science and Technology)
 - Big Data Era in sky and Earth Observation <http://bigskyearth.eu/>
 - Center for Astrostatistics (PennState) <http://astrostatistics.psu.edu/>

Participation aux organisations structurantes

- COSMOSTAT (J. L. Starck)
 - <http://www.cosmostat.org/>
- MADICS (<http://www.madics.fr/>)
 - Masses de Données, Informations et Connaissances en Sciences
 - Formations :
 - <http://www.madics.fr/reseaux/formation/liste-des-masters-du-domaine/>
- ARQUADS : Action de Recherche sur la Qualité des Données Scientifiques
 - <http://www.madics.fr/actions/actions-en-cours/arquads/>
- MAESTRO : Masse de données en Astrophysique :
 - <http://www.madics.fr/actions/actions-en-cours/maestro/>
 - Plateformes Type GALACTICA
 - Algorithmes de traitements BigData
 - Optimisation de traitement et de stockage



et une participation nationale

Formation

- ASTROINFO 2018
- June 25-29 - Marseille
- <https://astroinfo2018.sciencesconf.org/myspace/index>

- STAT4Astro
- every 2 years (next in 2019)
- <https://stat4astro2017.sciencesconf.org/>

- CosmoStat and Astronomical Data analysis
- <http://ada.cosmostat.org/>
- <http://cosmo21.cosmostat.org/>

- Astrostatistics
- <http://astrostatistics.psu.edu/datasets/index.html>



Vers une structuration AMU

- Pôles de Recherche Interdisciplinaires et Intersectoriels (PR2I)
 - PR2I Sciences et Technologies Avancées (P. Delaporte (LP3))
 - Bigdata – Machine Learning (C. Diaconu CPPM)
- Proposition d'Institut AMU regroupant les activités de recherche utilisant données scientifiques massives à l'AMU

